Chapter 10
Ontology Learning and the Humanities

Toby Burrows
The University of Western Australia, Australia

ABSTRACT
This chapter reviews the current state of play in the use of ontologies in the humanities, with best-practice examples from selected disciplines. It looks at the specific domain problems faced by the humanities, and examines the various approaches currently being employed to construct, maintain, and develop humanities ontologies. The application of ontology learning in the humanities is discussed by reviewing a range of research projects in different disciplines. The chapter concludes with an assessment of the future potential of ontology learning in the humanities, and an attempt to set out a research agenda for this field.

INTRODUCTION
The humanities are academic disciplines which study the nature of human life and experience. They are different from the natural and social sciences because they use methods which are mainly analytical, critical, or speculative. There are various different definitions of the scope of the humanities. According to the Australian Academy of the Humanities they cover the following disciplines: Archaeology; Asian Studies; Classical Studies; English; European Languages and Cultures; History; Linguistics; Philosophy, Religion and the History of Ideas; Cultural and Communication Studies; the Arts.¹

Ontological frameworks are central to the work of humanities researchers. This is because most humanities research involves either the analysis and definition of concepts or the categorization of individual phenomena into broader classes. Philosophy is the pre-eminent academic discipline which focuses on concepts, while the focus
Ontology Learning and the Humanities

of the disciplines of history and archaeology is largely on the categorization of specific instances (people, places, events, objects and so on). Other humanities disciplines rely on a mixture of these two approaches.

This paper reviews the current state of play in the use of ontologies in the humanities, with best-practice examples from selected disciplines. It looks at the specific domain problems faced by the humanities, and examines the various approaches currently being employed to construct, maintain and develop ontologies.

The application of ontology learning in the humanities is also examined, by reviewing a range of research projects in different disciplines. Areas discussed include the availability of text corpora and other sources of knowledge, and the use of text mining techniques and tools. The standards and tools used for expressing and developing ontologies are also covered.

The paper concludes with an assessment of the future potential of ontology learning in the humanities, and an attempt to set out a research agenda for this field. It also aims to identify areas where ontology learning is likely to prove most valuable and applicable.

Ontologies and the Humanities

The humanities are a particularly difficult area for the development and application of ontologies. The semantic context is complex and often ambiguous, and there are numerous existing vocabularies and taxonomies which often overlap. The multilingual nature of much humanities research has obvious implications for the use of ontologies. Not only are there different terms in different languages for the same concept; a term may be similar in appearance across different languages but have quite different meanings. Humanities research also often crosses different time periods, and the meanings of words and concepts tend to shift over time. Other worldviews and ways of categorizing the world – both current and past – also need to be factored in. Indigenous knowledge (the Australian aborigines or the North American Indians, for example) has a radically different hierarchy of classes, in addition to the obvious linguistic differences. A further area of complexity is the enormous proliferation of instances, particularly of people and places.

Nagypál (2005, pp. 208-209) discusses four specific difficulties:

• Time dependence: almost every instance is time-dependent, e.g. Strasbourg is part of France today, but has been part of Germany at various times in the past. This means that the relationship between an instance and its class is likely to be different at different historical periods.

• Uncertainty: the documentation relating to instances is frequently missing or contradictory, especially about dates. There are at least two different recorded birth dates for Joseph Stalin, for instance: 21 December 1879 in the official Soviet records, and 6 December 1878 in the church records. In another typical example, the paintings of the Dutch artist Vermeer cannot be dated precisely. This kind of uncertainty may affect the relationship between an instance and its class.

• Subjectivity: “most complex historical notions are vaguely defined or open to multiple interpretations, and thus can be interpreted subjectively, making them difficult to model conceptually.” Concepts like ‘the Enlightenment’ or ‘the Middle Ages’ do not have precise beginning and ending dates, for example, and it can be unclear whether a given temporal instance falls within the scope of a concept like this. Opinions and interpretations can differ considerably, even between experts.

• “Why” questions: historians tend to focus on this kind of question, and are interested in seeing facts in an explanatory context.